silicon systems™
INNOVATORS IN INTEGRATION

SSI 263A Phoneme Speech Synthesizer

Data Sheet

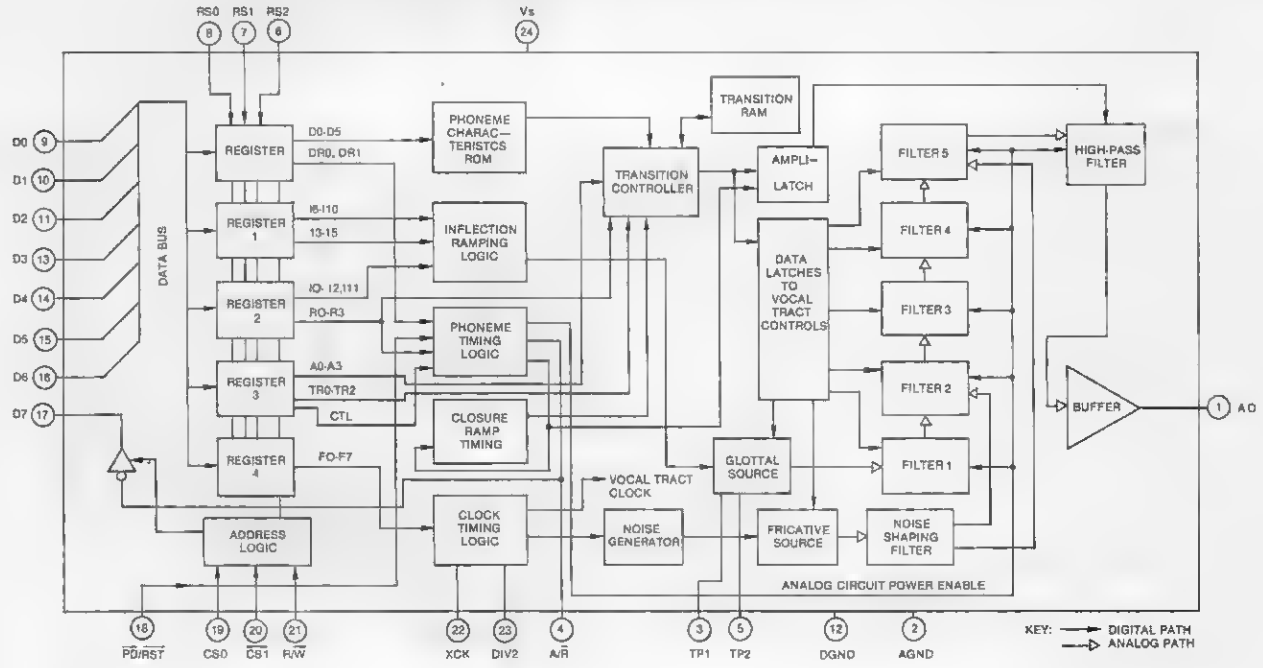
DESCRIPTION

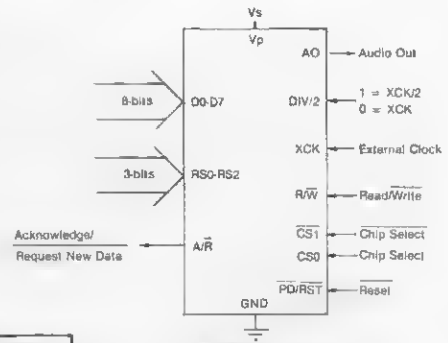
The SSI 263A is a versatile, high-quality, phoneme-based speech synthesizer circuit contained in a single monolithic CMOS Integrated circuit. It is designed to produce an audio output of unlimited vocabulary, music and sound effects at an extremely low data input rate.

Speech is synthesized by combining phonemes, the building blocks of speech, in an appropriate sequence. The SSI 263A contains five eight-bit registers that allow software control of speech rate, pitch, pitch movement rate, amplitude, articulation rate, vocal tract filter response, and phoneme selection and duration.

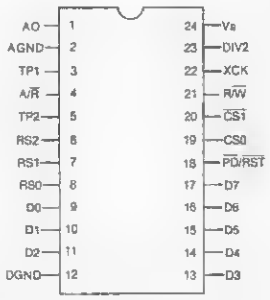
FEATURES

- **Single low-power CMOS integrated circuit**
- **5 Volt supply**
- **Extremely low data rate**
- **8-bit bus compatible with selectable handshaking modes**
- **Non-dedicated speech, ideal for text-to-speech programming**
- **Programmable and hard powerdown/reset mode**
- **Switched-capacitor-filter technology**





Signal Diagram



SSI 263A Pin Out
(Top View)

CAUTION: Use handling procedures necessary for a static sensitive component

SSI 263A Operation Description

This short description is intended to provide SSI 263A feature and capability information only. Refer to the SSI 263A USERS GUIDE for complete information on application and phonetic programming.

The Production of Speech

To produce different speech phonemes (sounds) the SSI 263A uses a model of the human vocal tract. Within the device this analog tract is modeled with five cascaded programmable low pass filter sections. The filter sections are programmed internally by a digital controller. Either a glottal (pitch) or a pseudo-random noise source is used to excite the vocal tract, depending on whether a voiced or non-voiced phoneme is selected. During speech production the phonemes will typically last between 25 and 100 mS.

The Speech Attribute Registers

Speech is produced by programming speech attribute (characteristic) data into five eight-bit registers. These internal registers allow selection of phonemes and speech characteristics. Refer to the Register Input Formats for the functional allocations.

Device Response to Attribute Register Data

The SSI 263A has two general classes of attribute data: "control" data (speech rate, filter frequency, phoneme articulation rate, phoneme duration, immediate inflection setting, and inflection movement rate) and "target" data (phoneme selection, audio amplitude, and transitioned inflection). The SSI 263A responds immediately upon loading "control" data; upon loading "target" data the device will begin to move towards that target at the prescribed transition rates. This fully internal linear transitioning between target values, done in a manner as is found in normal speech, is a key factor in reducing control data rate without sacrificing speech quality.

Attribute Register Writing

The eight bit data bus D7-D0 loads the particular attribute register selected by the three bit address bus RS2-RS0. To write the data, R/$\overline{W}$ (Read/Write), CS0 (Chip Select 0), and $\overline{CS1}$ pins must first be in the 0,1,0 state, respectively. The data is then written when at least one of these pins changes state. Refer to the Write Timing Diagram. Writing is accomplished by changing preferably CS0 or $\overline{CS1}$. Following device power up, nominal values should be loaded into the attribute registers as described below.

Approximate Data Transfer Rate

For speech production using the SSI 263A, the actual data rate depends on the amount of speech attribute manipulation. For example, the production of monotonic speech, where phoneme and duration are the only attribute manipulations, requires a data rate less than 100 bits-per-second. A higher data rate of about 500 bits-per-second is required for high quality speech due to the associated full attribute manipulation.

Selectable Operation Modes

The state of the Duration/Phoeme Register bits DR1 and DR0 determine the operating mode of the device when the Control bit (CTL) is changed from a logic one to a logic zero. The four modes of operation include choice of timing response between "frame" or "phoneme" timing (as explained below), transitioned or immediate inflection response, and setting the A/$\overline{R}$ (Acknowledge/Request Not) pin active or disabled. Refer to the Mode Selection Chart.

Phoneme Selection

The SSI 263A can produce the 64 phonemes listed on the Phoneme Chart. Bits P5-P0 are used for phoneme selection. The relative phoneme duration is set by bits DR1 and DR0.

Phoneme Articulation Adjustment

A particular phoneme is produced by the combination of vocal-tract low-pass filter settings, excitation source type, and source amplitude. When a new phoneme is selected, the device performs a linear transition to the new set of characteristics. The rate of this transition is controlled by the articulation setting, bits TR2-TR0. This rate is relative in that articulation is not affected by speech rate bits R3-R0. A typical articulation register setting is "5".

Programming Inflection (Pitch)

When the SSI 263A is in the mode of immediate inflection, bits I11-I0 provide immediate adjustment with seven octaves of pitch on an even tempered scale. With the device in the transitioned inflection mode, bits I10-I6 select the target pitch and bits I5-I3 determine the inflection rate of change. Bits I11, I2, I1, and I0 always provide immediate adjustment. A typical value used for speech production is 90Hz where:

$$\text{Inflection Frequency} = \frac{\text{XCK frequency}}{8 \times (4096 - I)}$$

I = decimal equivalent of Inflection Register setting

Filter Frequency Setting

Data bits FF7-FF0 set the clock frequency for the switched-capacitor vocal tract filters. This determines overall filter frequency response. Inflection pitch is not affected by these bits. Typically this is set to give a clock frequency of about 20KHz (see formula below), but may be manipulated to fine-tune speech quality or to change "voice type"; bass, baritone, etc.

$$\text{Filter Frequency} = \frac{\text{XCK frequency}}{2\,(256 - FF)}$$

FF = decimal equivalent to the Filter Frequency Register setting.

Speech Rate

Rate of speech is controlled by bits R3-R0, the Speech

Rate Register. In Frame Timing Mode new attribute data is requested at the end of a "frame" where:

$$\text{Frame Duration} = \frac{4096 \times (16-R)}{\text{XCK frequency}}$$

R = decimal equivalent of Rate Register setting

In the Phoneme Timing Mode the frame duration is modified by the phoneme duration bits DR1 and DR0 where:

Phoneme Duration = (Frame Duration) X (4-D)

D = decimal equivalent of Duration Register setting

All internal attribute transitioning is performed relative to the Speech Rate Register setting. Speech rate does not effect inflection or filter frequency. A typical rate setting is hexadecimal "A".

Amplitude Adjustment

The overall Audio Output level is set with register bits A3-A0. Since each phoneme has a preset amplitude relative to other phonemes, it is not necessary to program the amplitude of each phoneme; however, amplitude changes may be used to enhance the speech quality and add emphasis. Amplitude is transitioned linearly at rate dependent on the phoneme duration setting. A typical amplitude setting is hexadecimal "C".

Control Bit and Power Down Mode

Setting the Control bit (CTL) to a logic one puts the device into Power Down mode, a sort of "standby". This bit is also set high when the $\overline{PD}/\overline{RST}$ pin is brought low and also upon power up. The Power Down mode turns off the excitation sources and analog circuits to reduce power consumption, but maintains the present register settings. Upon a Control bit logic one-to-zero transition, the present settings of DR1 and DR0 determine the operation mode as described above.

Register Reading .

Device pin D7 becomes an output, as the inverted state of $A/\overline{R}$, when the device is put into Read ($R/\overline{W}$ is a logic 1 and the chip is selected, $\overline{CS1} = 0$, $CS0 = 1$). Refer to the Read Timing Diagram. The register address bits are ignored.

Time Base

Many different time bases may be utilized (see external clock input specifications). It is desirable to establish a stable crystal controlled time base from 800 to 1000KHz when DIV2 is set low, or twice the frequency when DIV2 is set high. A good time base can be easily accomplished with an inexpensive colorburst 3.5795 MHz crystal in conjunction with a divide-by-two circuit. The actual device timing and output frequencies are directly related to the time base frequency used.

Microprocessor Interfacing

Either the $A/\overline{R}$ line, or D7 as an output, are used as an interrupt to indicate when the duration of a frame or phoneme has been exceeded. No detectable degradation to speech quality results when several milliseconds occur between data request and load.

PHONEME CHART

Hex Code*	Phoneme Symbol	Example Word (or Usage)
00	PA	(pause)
01	E	MEET
02	E1	BENT
03	Y	BEFORE
04	YI	YEAR
05	AY	PLEASE
06	IE	ANY
07	I	SIX
08	A	MADE
09	AI	CARE
0A	EH	NEST
0B	EH1	BELT
0C	AE	DAD
0D	AE1	AFTER
0E	AH	GOT
0F	AH1	FATHER
10	AW	OFFICE
11	O	STORE
12	OU	BOAT
13	OO	LOOK
14	IU	YOU
15	IU1	COULD
16	U	TUNE
17	U1	CARTOON
18	UH	WONDER
19	UH1	LOVE
1A	UH2	WHAT
1B	UH3	NUT
1C	ER	BIRD
1D	R	ROOF
1E	R1	RUG
1F	R2	MUTTER (German)
20	L	LIFT
21	L1	PLAY
22	LF	FALL (final)
23	W	WATER
24	B	BAG
25	D	PAID
26	KV	TAG (glottal stop)
27	P	PEN
28	T	TART
29	K	KIT
2A	HV	(hold vocal)
2B	HVC	(hold vocal closure)
2C	HF	HEART
2D	HFC	(hold fricative closure)
2E	HN	(hold nasal)
2F	Z	ZERO
30	S	SAME
31	J	MEASURE
32	SCH	SHIP
33	V	VERY
34	F	FOUR
35	THV	THERE
36	TH	WITH
37	M	MORE
38	N	NINE
39	NG	RANG
3A	:A	MARCHEN (German)
3B	:OH	LOWE (French)
3C	:U	FUNF (German)
3D	:UH	MENU (French)
3E	E2	BITTE (German)
3F	LB	LUBE

*Note — Hex codes shown with DR0, DR1 = 0 (longest Duration)

PIN ASSIGNMENT DESCRIPTIONS

Pin No.	Symbol	Active Levet	Description
1	AO		Analog Audio Output biased @ VDD/2 requires an external audio amp for speaker drive
2	AGND		Analog Ground
3	TP1		Do not use
4	A/$\overline{R}$		Acknowledge/Request Not — open collector output changes from high to low level after phoneme is generated. May be used as an Interrupt request for new phoneme data. (See Pin 17 also.)
5	TP2		Do not use
6	RS2		Register Select Input – used to select one of five internal registers in conjunction with RS1 and RS0
7	RS1		Register Select (See pin 6)
8	RS0		Register Select (See pin 6)
9	D0		LSB of 8-bit data bus — input only
10	D1		Data Input (only)
11	D2		Data Input (only)
12	DGND		Digital Ground
13	D3		Data Input (only)

Pin No.	Symbol	Active Levet	Description
14	D4		Data Input (only)
15	D5		Data Input (only)
16	D6		Data Input (only)
17	D7		MSB of 8-bit data bus. Bi-directional, inverse of pin 4 when read is high
18	PD/$\overline{RST}$	Low	Power Down Control Input — Silences audio output and retains DC bias without disturbing register contents. Disables A/$\overline{R}$ output.
19	CS0	High	Chip Select Input
20	$\overline{CS1}$	Low	Chip Select Input
21	R/$\overline{W}$		Read/Write Control Input — Write is active low for loading internal registers. Read is active high but enables D7 only.
22	XCK		Clock Input (≃1 or 2 MHz)
23	DIV2	High	Clock Divide by Two — used when external clock is ≃ 2 MHz
24	VDD		Positive Voltage Supply

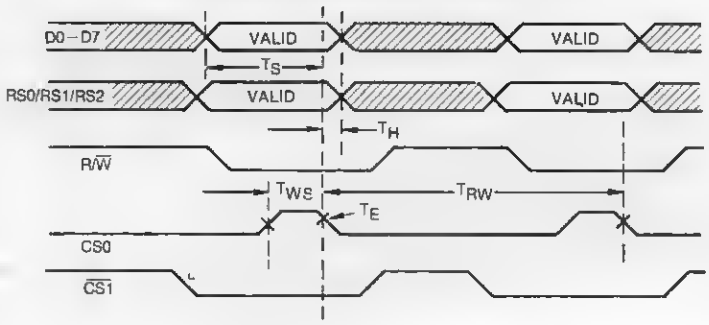
REGISTER INPUT FORMATS

Register Address			Register Name	Bus Input Bit Position							
RS2	RS1	RS0		D7	D6	D5	D4	D3	D2	D1	D0
LO	LO	LO	Duration/Phoneme (DR/P)	DR1	DR0	P5	P4	P3	P2	P1	P0
LO	LO	HI	Inflection (I)	I10	I9	I8	I7	I6	I5	I4	I3
LO	HI	LO	Rate/Inflection (R/I)	R3	R2	R1	R0	I11	I2	I1	I0
LO	HI	HI	Control/Articulation/Amplitude (C/A/A)	CTL	T2	T1	T0	A3	A2	A1	A0
HI	X	X	Filter Frequency (F)	F7	F6	F5	F4	F3	F2	F1	F0

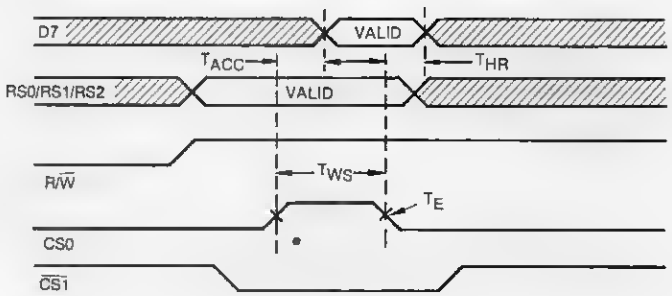
DR1, DR0 . . Define the phoneme duration.
P5 ←P0 . . . Address the phoneme required.
I11←I0 . . . . Define inflection target frequencies and rate of change.
R3 ←R0 . . . Define the rate or speed of speech.
CTL . . . . . Define the mode of A/$\overline{R}$ response in conjunction with DR1 and DR0. Also directly set by PD/$\overline{RST}$.

T2←T0 . . . . Define the rate of movement of the formant position for articulation purposes.
A3←A0 . . . Define the amplitude of the output audio.
F7←F0 . . . Define the frequency of all vocal tract filters.

WRITE TIMING DIAGRAM



READ TIMING DIAGRAM



*Valid data latched on first rise or fall of R/W, CS0 or CS1 into inactive.

Timing Characteristics (V_{DD} = 4.5 to 5.5 Volts, TA = -40 to +85 deg. C)

Item	Symbol	Limits		Units.
		Min.	Max.	
Data Setup Time	TS	120**		nsec
Data Hold Time	TH	10**		nsec
Strobe Width	TWS	200		nsec
Read/Write Cycle Time	TRW	2.25*		μsec
Rise/Fall Time	TE		100	nsec
D7 Output Access Time	TACC		180	nsec
D7 Output Hold Time	THR		180	nsec

Notes: * Based on color burst frequency.
** Timing relative to deselect by either CS0, CS1, or R/W changing.

MODE SELECTION CHART

DR1	DR0	'CTL' BIT	Function
HI	HI	HI→LO	A/$\bar{R}$ active; phoneme timing response; transitioned inflection (most commonly used mode)
Hi	LO	HI→LO	A/$\bar{R}$ active; phoneme timing response; immediate inflection
LO	HI	HI→LO	A/$\bar{R}$ active; frame timing response; immediate inflection
LO	LO	HI→LO	Disables A/$\bar{R}$ output only; does not change previous A/$\bar{R}$ response

ABSOLUTE MAXIMUM RATINGS

Item	Symbol	Limit	Units
Supply Voltage	$V_{DD}-V_{SS}$	7.0	V
Input Voltage	V_{IN}	–0.5 to V_{DD} + 0.5	V
D.C. Current at Inputs	I_{INM}	± 1.0	mA
Storage Temperature	T_S	–55 to + 125	°C
Operating Temperature	T_A	–40 to + 85	°C
Power Dissipation	P_d	500	mW

Electrical Characteristics Unless otherwise specified, $4.5 \leq V_{DD} \leq 5.5$; -40 deg. C $\leq$ TA $\leq$ 85 deg. C;
1.50MHz $\leq$ XCK frequency $\leq$ 2.0MHz, when XCK/2 = logic 1 or
0.75MHz $\leq$ XCK frequency $\leq$ 1.0MHz, when XCK/2 = logic 0

Description	Conditions	Min.	Typ.	Max.	Units
POWER SUPPLY					
Supply Current	$\overline{PD}/\overline{RST} = 1$, CTL = 0		8	20	mA
Supply Current	$\overline{PD}/\overline{RST} = 0$, CTL = 1		7	18	mA
AUDIO OUTPUT					
Output Level	AW phoneme RL = 50Kohm to GND through 1μF cap.	0.28VDD	0.37VDD	0.50VDD	Vpp
DC Output Offset		0.5VDD	0.6VDD	0.7VDD	V
Resistive Loading	AC coupled to AO to GND	10			Kohm
Capacitive Loading	To GND to ensure Stable A			100	pF

Description	Conditions	Symbol	Min	Typ	Max	Units
BUS CONTROL INPUTS, DATA INPUTS (RS0, RS1, RS2, CS0, $\overline{CS1}$, D0-D7 $\overline{PD}/\overline{RST}$)						
Input High Voltage		V_{IH}	$V_{SS}+2.4$		$V_{DD}+0.3$	VDC
Input Low Voltage		V_{IL}	-0.3		$+0.8$	VDC
Input Leakage Current	$V_{IN} = 0$ to V_{DD}	I_{IN}			5	μA
Input Capacitance	$V_{IN} = 0$ $T_A = 25°C$ measured at f = 1.0MHz	C_{IN}			10	pF
Input Capacitance, D7 Input		$C_{IN(D7)}$			20	pF
Input Current, D7 In TRI-State "OFF" State	V_{IN} = 0.4 to 2.4 V	$I_{IN(TS)}$		2.0	5.0	μA
D7 OUTPUT						
D7 Output Low Voltage	$I_{Load} = 0.4$ mA into D7	$V_{OL(D7)}$			0.4	VDC
D7 Output High Voltage	$I_{Load} = 205\mu$A out of D7	$V_{OH(D7)}$		V_{DD}-2.0		VDC
A/$\overline{R}$ OUTPUT						
Output Low Voltage	$I_L = 3.2$ mA into A/$\overline{R}$	$I_{OL(A/\overline{R})}$			0.4	VDC
Output High Leakage Current	$V_{Out} = 0.0$ to V_{DD}	$I_{L(A/\overline{R})}$			10	μA
Output Capacitance	$V_{Out} = 0$ VDC $T_{AMB} = 25°C$ f = 1.0 MHz	$C_{Out(A/\overline{R})}$		15	pF	
DIV2 INPUT						
Input Low Voltage		$V_{IL(DIV2)}$	-0.3		.2 V_{DD}	V
Input High Voltage		$V_{IH(DIV2)}$	.8V_{DD}		$V_{DD}+0.3$	V
Input Leakage	$V_{IN} = 0$ to V_{DD}				5	μA

Description	Conditions	Symbol	Min.	Typ.	Max.	Units.
XCLK						
Input Low Voltage		$V_{IH}(IC)$	—0.3		+0.8	V
Input High Voltage		$V_{IH}(IC)$	2.4		V_{DD} + 0.3	V
Input Current	V_{IN} = 0.0 to V_{DD}	$I_{IN}(C)$			5	μA
Input Capacitance		$C_{IN}(C)$			10	pF
Duty Cycle		D(XCLK)	0.4		0.6	—

TYPICAL MICROPROCESSOR IMPLEMENTATION